Re-identification with multiple source-cameras

Syed Fahad Tahir*†, Andrea Cavallaro* and Bernhard Rinner†
*Centre for Intelligent Sensing, Queen Mary University of London, London, UK
†Institute of Networked and Embedded Systems, Alpen-Adria-Universität, Klagenfurt, Austria
Email: {s.fahad.tahir, andrea.cavallaro}@eecs.qmul.ac.uk, bernhard.rinner@aau.at

***Abstract*—Target re-identification approaches generally perform association between camera pairs only. However, in a multi-camera system many-to-one camera associations are needed when targets transit from multiple source-cameras to a destination-camera. To address this problem, we propose a person re-identification approach that generates camera-invariant object matching scores, which are based on re-identification score variations in multiple camera pairs. Each camera pair is represented with two parametric distribution models obtained by curve fitting on intra-class and inter-class target matching scores. These two models are combined to generate the likelihood of a correct match between a new target in the destination-camera and those in all source-cameras. We show the improvement in the performance of the proposed re-identification approach compared to existing pairwise approaches on two publicly available datasets.**

Keywords*—Camera network, multiple source-cameras, many-to-one association, re-identification, distribution estimation, probability density function, appearance information.

I. Introduction

Re-identification aims at the recognition of the same object that appears over time in different cameras with non-overlapping fields of view [1], [2]. Re-identification is usually performed in a pairwise manner assuming that the source-camera (i.e. the camera whose field of view was left by the object) is known [3], [4], [5], [6], [7], [8]. However, in real-world camera networks, an object in a destination-camera could have been generated by multiple candidate source-cameras.

In this paper, we propose a re-identification approach for the case when a camera detects objects that can come from different source-cameras. We estimate the distributions of matching scores obtained by association of objects in each camera pair. We apply a data over-sampling technique to balance the number of obtained matching scores between the same persons and different people in a camera pair. Using the distributions of these scores, we measure probabilities of a correct match from the objects detected in a group of source-cameras. We generate the camera-invariant matching scores/likelihoods by maximizing the differences between the probabilities of being the same or different objects. Association is then performed by optimal assignment using the Hungarian algorithm.

The paper is organized as follows. In Sec. II, we discuss existing feature sets and pairwise association approaches. Sec. III describes the steps involved in the proposed approach. In Sec. IV, the approach is validated and compared with existing re-identification approaches using two multi-camera datasets. Finally, Sec. V draws conclusions and discusses the future work.

II. Previous work

In this section we briefly review the features and the association methods that are used in the re-identification literature.

Features include color, texture and shape of an object [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12]. Features to be used for re-identification need to present a certain level of invariance to account for the pose, illumination and size variations that are likely to occur in different camera views. Appearance information can be extracted from a single instance of an object [4], [5], [10] or from multiple instances of the same object in the same camera view in order to generate a more robust representation [13]. Histograms of color-channels such as RGB [7], [9], [14], HSV [6], and YUV [15] can be used as *color* features. Gabor and Schmid filters are applied to one image channel and the histograms of convolved images are used as *texture* information [3], [5]. Local binary patterns [16] and feature point descriptors like SIFT [13] are also applied to extract textures. The *shape* of an object can be preserved by region covariance descriptor [17] and histograms of oriented gradients [13]. Concatenation of histograms is also common [3], [4], [5]. Multiple features can be combined to improve inter-object discrimination. Moreover, features can be selected using the Attribute-Sensitive Feature Importance (ASFI) method [18] or the Cost-and-Performance-Effective (CoPE) feature-selection method [10].

Association for re-identification can be performed by feature correlation [11], direct distance minimization (DDM) and learning approaches. *DDM* measures the similarity (distance/score) between the features in a camera pair using for example the Bhattacharyya Distance [6], [19], the Euclidean distance, Kullback-Leibler Distance [7], [15], the sum of absolute [20] or quadratic [21] distances. DDM is less robust to illumination changes, which can be compensated by the cross-camera color *calibration* that maps the color features between a camera pair before the association [15], [19]. Feature transformation and projection [22] may also be applied. In addition to AdaBoost [5] and rankSVM [4], another *learning* approach used in pairwise re-identification is Probabilistic Relative Distance Comparison (PRDC) [3], which maximizes the probability of correct matches and minimizes that of incorrect matches by learning their relative distances in a camera pair.

S.F. Tahir was supported by the Erasmus Mundus Joint Doctorate in Interactive and Cognitive Environments, which is funded by the Education, Audiovisual & Culture Executive Agency (FPA n° 2010-0012).

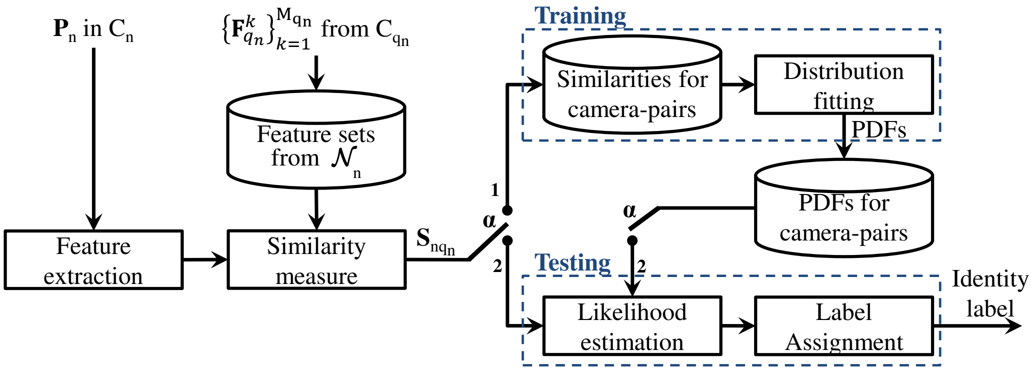

Fig. 1. Block diagram of the proposed re-dentification approach for multiple source-cameras. Switch $\alpha = 1$ is Training and $\alpha = 2$ is Testing. $\mathcal{N}_n$ is the set of neighboring cameras of C_n.

III. ASSOCIATION WITH MULTIPLE SOURCE-CAMERAS

A. Overview

Let $\mathbf{C} = \{C_n\}_{n=1}^{N}$ be a network of N cameras. Let C_{q_n} be defined as neighbor of C_n if an object leaving C_{q_n} can enter C_n without passing through any other cameras. Each C_n has $\hat{N}_n$ neighboring cameras $\mathcal{N}_n = \{C_{q_n}\}_{q_n=1}^{\hat{N}_n}$, where $q_n \neq n$, $\hat{N}_n \leq N$ and $\mathcal{N}_n \subseteq \mathbf{C}$.

Let a set $\mathbf{P}_n = \{P_n^m\}_{m=1}^{M_n}$ of M_n objects be detected in C_n. For each P_n^m we aim to identify its other instance $P_{q_n}^k$ detected in an unspecified camera in $\mathcal{N}_n$, where $k = 1 \cdots M_{q_n}$ and M_{q_n} objects are detected in C_{q_n}. This can be cast as a content-based information retrieval problem, where the most relevant matches are retrieved/ranked for a given query P_n^m from multiple information sources ($\mathcal{N}_n$).

In order to perform many-to-one camera association for retrieving the correct match, we aim to generate camera-invariant matching scores from the matching model of each camera pair (C_n, C_{q_n}) obtained by exploiting similarities between objects.

The proposed association approach (Fig. 1) estimates the likelihood of a correct match in the case of objects coming from multiple source-cameras. If M_{q_n} objects are generated by each source-camera C_{q_n} that go to C_n, the the number of objects M_n detected in C_n is

$$M_n = \sum_{q_n=1}^{\hat{N}_n} M_{q_n}. \tag{1}$$

From each detected object in each camera we extract the feature set $\mathbf{F}_n^m = \{\mathbf{f}_n^{mr}\}_{r=1}^{R}$ containing R features. We apply the pairwise association approaches on $\mathbf{F}_n^m$ and the obtained feature sets $\{\mathbf{F}_{q_n}^k\}_{k=1}^{M_{q_n}}$ from C_{q_n} to get M_{q_n} similarity scores $\{S_{nq_n}^{mk}\}_{k=1}^{M_{q_n}}$ for P_n^m.

B. Training

In the training phase, we get a set of similarity scores $\mathbf{S}_{nq_n}$ between M_{q_n} objects detected in a camera pair (C_n, C_{q_n})

$$\mathbf{S}_{nq_n} = \left\{ \left\{ S_{nq_n}^{mk} \right\}_{m=1}^{M_{q_n}} \right\}_{k=1}^{M_{q_n}}. \tag{2}$$

$\mathbf{S}_{nq_n}$ contains $M_{q_n} \times M_{q_n}$ elements. We divide $\mathbf{S}_{nq_n}$ in two subsets $\mathbf{S}_{nq_n}^{+}$ and $\mathbf{S}_{nq_n}^{-}$, which respectively contain the similarity scores for the same objects $\mathcal{S}$ and the different objects $\overline{\mathcal{S}}$ in a camera pair. Training for re-identification suffers from under-sampling due to the availability of few object-images and many pose and illumination changes [3], [10]. In addition, $|\mathbf{S}_{nq_n}^{+}| << |\mathbf{S}_{nq_n}^{-}|$ results in an unbalanced class problem ($|.|$ is the cardinality of a set). In order to compensate for the under-sampled and unbalanced data, we include more related-samples, generated by applying the oversampling technique SMOTE [23] on the scores in $\mathbf{S}_{nq_n}^{+}$ and $\mathbf{S}_{nq_n}^{-}$. SMOTE measures the difference between the sample and its nearest neighbor(s), and the difference is added to the sample under consideration to generate similar synthetic examples. Next, we normalize the histograms of $\mathbf{S}_{nq_n}^{+}$ and $\mathbf{S}_{nq_n}^{-}$ to obtain their corresponding Probability Density Functions (PDFs). We characterize the PDFs by fitting the existing parametric distribution models [24]. The two models $T_{nq_n}^{+}$ and $T_{nq_n}^{-}$ nearest to the PDFs of $\mathcal{S}$ and $\overline{\mathcal{S}}$ are selected by applying Bayesian Information Criterion [25] (Fig. 2).

The performance of the approach can be improved by increasing the number of objects detected in each camera pair and available for training, which needs to be performed only once during the camera network setup. If M_{q_n} objects are required for the training of a camera-pair (C_n, C_{q_n}), the addition of a new destination-camera C_n to N cameras of the network would require $M_{q_n} \times \hat{N}_n$ objects that move from $\hat{N}_n$ source-cameras to the new camera. In the worst-case scenario when every destination camera has $N-1$ source cameras, the training required for the N cameras network is increased by $N(N-1)/2$.

C. Testing

In the testing phase, a new object P_n^m is detected in C_n and feature sets are received from the set of neighboring cameras $\mathcal{N}_n$. We measure the similarity score $S_{nq_n}^{mk}$ using $\mathbf{F}_n^m$ and each obtained feature set $\mathbf{F}_{q_n}^k$ from C_{q_n}. Next for the given $S_{nq_n}^{mk}$, we measure the probabilities $p(\mathcal{S}|S_{nq_n}^{mk})$ and $p(\overline{\mathcal{S}}|S_{nq_n}^{mk})$ of P_n^m and $P_{q_n}^k$ to be the two instances of the same or different objects, respectively, given as

$$\begin{aligned} p(\mathcal{S}|S_{nq_n}^{mk}) &= G(S_{nq_n}^{mk}, T_{nq_n}^{+}), \\ p(\overline{\mathcal{S}}|S_{nq_n}^{mk}) &= G(S_{nq_n}^{mk}, T_{nq_n}^{-}), \end{aligned} \tag{3}$$

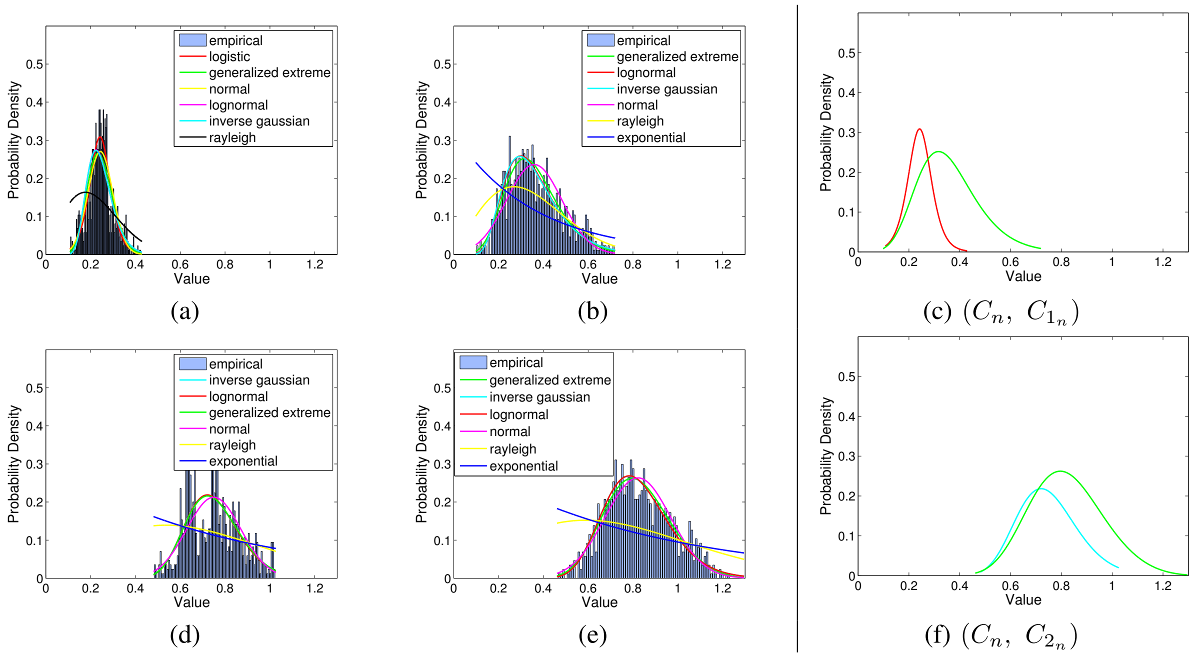

Fig. 2. An example of distributions of matching distances between objects detected in C_n and one of its two neighboring cameras (top row) C_{1_n} and (bottom row) C_{2_n}. Legends are sorted in the order of nearest distribution to the data identified using Bayesian information criterion. Distances are between (a,d) same and (b,e) different objects in a camera pair. (c,f) The two selected distribution models for each camera pair.

where $G(.,.)$ takes as input the similarity score $S^{mk}_{nq_n}$ and one of the selected models from $T^{+}_{nq_n}$ and $T^{-}_{nq_n}$ of the corresponding camera pair (C_n, C_{q_n}) and returns the probabilities of being two objects the same or different. Since the two probabilities are independent, we can combine them to maximize the likelihood $L^{mk}_{nq_n}$ of P^m_n and $P^k_{q_n}$ to be a correct match as

$$L^{mk}_{nq_n} = p(\mathcal{S}|S^{mk}_{nq_n}) - p(\overline{\mathcal{S}}|S^{mk}_{nq_n}). \quad (4)$$

$L^{mk}_{nq_n} \in [-1, 1]$. The larger $L^{mk}_{nq_n}$, the higher the probability for the pair to be a correct match. For each P^m_n, we get M_n likelihoods from all the objects detected in $\mathcal{N}_n$ which results in the likelihood matrix $\mathbf{L}_n$ for the set of $\mathbf{P}_n$ objects given as

$$\mathbf{L}_n = \left[\left[\left[L^{mk}_{nq_n}\right]^{M_n}_{m=1}\right]^{M_{q_n}}_{k=1}\right]^{\hat{N}_n}_{q_n=1}. \quad (5)$$

$L^{mk}_{nq_n}$ is camera-invariant due to the individual modeling of distributions of similarity scores in each camera pair, therefore the assignment of P^m_n to $P^k_{q_n}$ coming from any C_{q_n} becomes possible. Finally, we select the correct from the obtained likelihoods $\mathbf{L}_n$ by optimal assignment using the Hungarian algorithm [12]. The algorithm performs group assignment without repetition such that the summation of likelihoods between all the assigned pairs in a group remains minimum.

IV. Evaluation and Discussions

We evaluate the proposed approach with learning, probabilistic and DDM based pairwise association approaches: PRDC [3], RankSVM [4], ASFI [26] and Bhattacharyya distance, using rank-ROC curves. Unlike CMC [5], rank-ROC curves explicitly show at each rank the false positive rate (FPR) [1-specificity] along with the true positive rate (TPR) [sensitivity].

Single images of persons per camera are manually extracted. We extract color and texture features. Each feature is a 16-bin histogram of a color channel or a filtered image, extracted from each of the 6 horizontal stripes of the person image. We use eight color channels (R, G, B, H, S, Y, Cb, Cr) from RGB, HSV, and YCbCr color spaces, and for texture, eight Gabor and 13 Schmid filters are applied on the Y channel of the image as in [5], [3], [18], [10].

We also apply the proposed association approach with the features selected by CoPE [10]. CoPE returns a list $\mathbf{Y}_{nq_n}$ of selected features for each camera pair such that P^m_n and $P^k_{q_n}$ are represented by $\mathbf{F}^m_n = \{\mathbf{f}^{m\hat{r}}_n\}_{\hat{r}\in\mathbf{Y}_{nq_n}}$ and $\mathbf{F}^k_{q_n} = \{\mathbf{f}^{k\hat{r}}_{q_n}\}_{\hat{r}\in\mathbf{Y}_{nq_n}}$, respectively. Using each selected feature $\mathbf{f}^{\hat{r}}$, we apply DDM (Bhattacharyya) to obtain a similarity score matrix $\mathbf{S}^{\hat{r}}_{nq_n}$ (Eq. 2) and estimate $T^{(\hat{r})+}_{nq_n}$ and $T^{(\hat{r})-}_{nq_n}$, for the camera pair $(C_n,\ C_{nq_n})$. For the new objects P^m_n and $P^k_{q_n}$, and each selected feature, we measure the likelihood $L^{mk(\hat{r})}_{nq_n}$ of a match (Eq. 4) and obtain a combined likelihood score $L^{mk}_{nq_n}$ for CoPE as

$$L^{mk}_{nq_n} = \frac{\sum\limits^{\hat{r}\in\mathbf{Y}_{nq_n}} L^{mk(\hat{r})}_{nq_n}}{|\mathbf{Y}_{nq_n}|}. \quad (6)$$

Finally, we obtain the matrix $\mathbf{L}_n$ (Eq. 5) for assignments.

We use two publicly available person datasets: the WARD dataset [13] and the Torch dataset [12]. *WARD* contains 70 persons from three non-overlapping fixed-cameras with the challenges of illumination changes, and variations in pose and size. *Torch* contains 50 persons from five hand-held smartphone cameras representing an outdoor crowd scene with additional challenges of occlusions, occasional jitters and blurring. We assume that the person detection problem is solved [27].

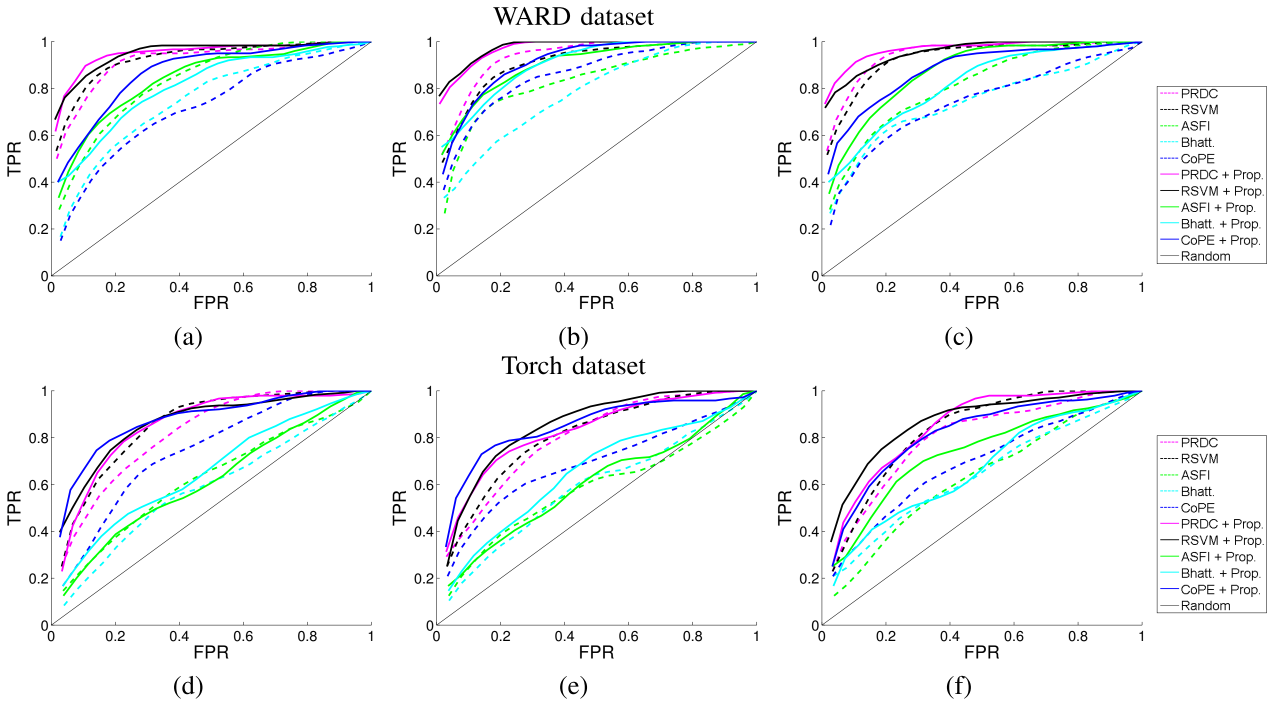

Fig. 3. Ranked ROC curves for re-identification in three-camera settings using the existing approaches: PRDC [3], ASFI [26], rankSVM [4] and CoPE [10] compared with the proposed association approach using, (a-c) WARD dataset [13] ($|\mathbf{P}_n| = 30$), and (d-f) Torch dataset [12] ($|\mathbf{P}_n| = 24$). Each camera detects persons that come from the other two as source-cameras such that persons appear in (a, d) C_1 from C_2 and C_3, (b, e) C_2 from C_1 and C_3, and (c, f) C_3 from C_1 and C_2. Key: Bhatt. - Bhattacharyya distance, Prop. - Proposed approach.

TABLE I. EXPERIMENTAL SETUP USED FOR THE EVALUATION.

Datasets	network size ($\|\mathbf{C}\|$)	Persons in C_n (M_n)	Neighboring cameras ($\|\mathcal{N}_n\|$)	Persons in C_{q_n} (M_{q_n})	FPs
WARD	2	30	1	30	-
	3	30	2	15	-
	3	38	2	15	8
Torch	2	24	1	24	-
	3	24	2	12	-
	3	30	2	12	6
	4	24	3	8	-
	5	24	4	6	-

We apply two-fold cross validation such that half of the dataset is used for training [Table I]. The number of persons detected in $\mathcal{N}_n$ are fixed to 30 in WARD, and 24 in the Torch dataset. We also perform the experiments with 25% added false positives (FPs), i.e. persons detected in C_n that do not appear in C_{q_n}[1]. Finally, we analyze the changes in re-identification rate by increasing the number of cameras.

Fig. 3 shows the ROC curves for re-identification in three-camera settings for WARD and the Torch datasets. Due to the likelihood measure in Eq. 4, the performances of the existing pairwise methods are improved by the proposed association approach. In WARD dataset, the less illumination changes resulting in more inter-camera similarities make it challenging to learn differences between camera pairs. The proposed approach increases TPR of PRDC and RSVM from the range between 0.25 and 0.55, upto 0.75 in the start of the curves. In Torch dataset, the re-identification rate is less compared to the WARD for all approaches due to the additional challenges of occlusions and blur; however, compared to the existing approaches, improvement in the re-identification rate can be observed by the proposed approach. CoPE using the proposed approach shows the highest improvement in the performance due to the likelihood estimation at the feature level (Eq. 6), while ASFI and DDM remain the least in both pairwise and three-camera settings.

Fig. 4 shows the ROC curves for re-identification with added FPs in three-camera settings (8 in WARD and 6 in Torch). In both datasets the re-identification rate is improved with the proposed association approach. In WARD, TPR of the proposed approach with PRDC and RSVM remain higher starting at 0.3 and reach to 1 at 80% of the FPR. In the Torch dataset, due to the more challenging settings, ASFI and Bhattacharyya do not perform well; however, the learning methods RSVM and PRDC with the proposed approach remain least effected and show improvement in TPR. CoPE shows the highest rate of improvement in the re-identification with the proposed association approach.

Finally, we analyze using the AUC of ROC how the re-identification performance varies as the number of cameras increases (Fig. 5). We use five cameras from the Torch dataset. Keeping the total number of persons detected in C_n fixed to 24, the experiments are performed for all combinations containing one, two, three and four source-cameras (Table I). The proposed approach improves the re-identification performance for all combinations of cameras in the network. As the number of cameras increases the performance decreases gradually;

[1] Note that while we do not perform experiments with false negative detection (i.e. persons detected in C_{q_n} that do not appear in C_n) because of the limited size of the dataset; using the optimal assignment used in the proposed approach, we would expect that the influence of false negatives would be comparable to that of additional false positives in terms of re-identification performance.

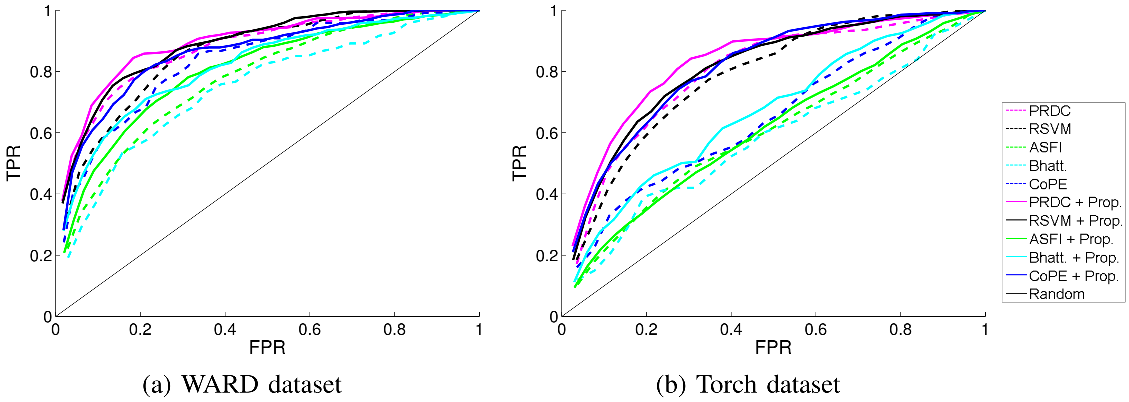

Fig. 4. Ranked ROC curves for re-identification in three-camera settings with 25% additional persons (FPs) using the existing approaches: PRDC [3], ASFI [26], rankSVM [4] and CoPE [10] compared with the proposed association approach on (a) WARD [13] and (b) Torch dataset [12]. Key: Bhatt. - Bhattacharyya distance, Prop. - Proposed approach.

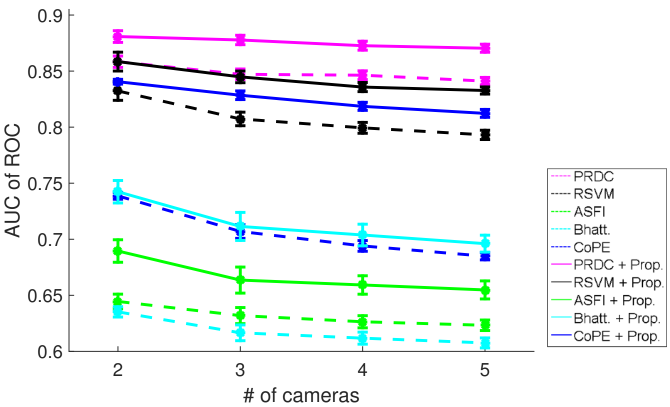

Fig. 5. AUC of ROC for re-identification as the number of cameras are increased from two to five in Torch dataset. Key: Bhatt. - Bhattacharyya distance, Prop. - Proposed approach.

however the rate of decrease in performance is relatively small, especially when the proposed association approach is applied with PRDC, RankSVM and CoPE. DDM shows the best improvement with the proposed approach.

V. Conclusions

We proposed an association approach to extend pairwise re-identification methods to multiple cameras. The proposed approach estimates the likelihood of a correct match in a camera network from the similarity scores in camera pairs. We showed that the proposed approach can improve the re-identification rate by 20% on two datasets, while the degradation in the performance as the number of cameras increases is smaller than for existing approaches.

As future work, we aim to extend the validation to a larger dataset and to represent each camera pair with a single mixture model of distributions and to represent multiple camera pairs with same models to improve scalability

References

[1] R. Vezzani, D. Baltieri, and R. Cucchiara, "People re-identification in surveillance and forensics: a survey," *ACM Computing Surveys*, vol. 46, no. 2, pp. 1–37, 2013.

[2] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: problem overview and current approaches," *Journal of Ambient Intelligence and Humanized Computing*, vol. 2, no. 2, pp. 127–151, 2011.

[3] W.-S. Zheng, S. Gong, and T. Xiang, "Re-identification by relative distance comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 653–668, 2013.

[4] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by support vector ranking," in *Proc. BMVC*, Aberystwyth, UK, Aug 2010.

[5] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in *Proc. ECCV*, Marseille, France, Oct 2008.

[6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Proc. CVPR*, San Francisco, USA, Jun 2010.

[7] G. Berdugo, O. Soceanu, Y. Moshe, D. Rudoy, and I. Dvir, "Object re-identification in real world scenarios across multiple non-overlapping cameras," in *Proc. EUSIPCO*, Aalborg, Denmark, Aug 2010.

[8] R. Mazzon, S. F. Tahir, and A. Cavallaro, "Person re-identification in crowd," *Pattern Recognition Letters*, vol. 33, no. 14, pp. 1828–1837, 2012.

[9] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views," *Computer Vision and Image Understanding*, vol. 109, no. 2, pp. 146–162, Feb 2008.

[10] S. F. Tahir and A. Cavallaro, "Cost-effective features for re-identification in camera networks," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 8, pp. 1–13, 2014.

[11] C.-H. Kuo, C. Huang, and R. Nevatia, "Inter-camera association of multi-target tracks by on-line learned appearance affinity models," in *Proc. ECCV*, Crete, Greece, Sep 2010.

[12] S. F. Tahir and A. Cavallaro, "Low-cost multi-cameras object matching," in *Proc. IEEE ICASSP*, Florence, Italy, May 2014.

[13] N. Martinel, C. Micheloni, and C. Piciarelli, "Distributed signature fusion for person re-identification," in *Proc. ACM/IEEE ICDSC*, Oct 2012.

[14] Y. Cheng, W. Zhou, Y. Wang, C. Zhao, and S. Zhang, "Multi-camera-based object handoff using decision-level fusion," in *Proc. IEEE CISP*, Tianjin, China, Oct 2009.

[15] K. Jeong and C. Jaynes, "Object matching in disjoint cameras using a colour transfer approach," *Springer Journal of Machine Vision and Applications*, vol. 19, no. 5, pp. 88–96, 2008.

[16] G. Lian, J.-H. Lai, C. Suen, and P. Chen, "Matching of tracked pedestrians across disjoint camera views using ci-dlbp," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 7, pp. 1087 –1099, 2012.

[17] O. Tuzel, F. Porikli, and P. Meer, “Region covariance: A fast descriptor for detection and classification,” in *Proc. ECCV*, Graz, Austria, May 2006, pp. 589–600.

[18] C. Liu, S. Gong, and C. C. Loy, “On-the-fly feature importance mining for person re-identification,” *Pattern Recognition*, vol. 47, no. 4, pp. 1602–1615, 2014.

[19] B. Prosser, S. Gong, and T. Xiang, “Multi-camera matching using bi-directional cumulative brightness transfer functions,” in *Proc. BMVC*, Leeds, UK, Sep 2008.

[20] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, “Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences,” in *Proc. ACM/IEEE ICDSC*, California, USA, Sep 2008.

[21] I. Oliveira and J. Luiz, “People re-identification in a camera network,” in *Proc. IEEE DASC*, Chengdu, China, Sep 2009.

[22] Y. Wang, R. Hu, C. Liang, C. Zhang, and Q. Leng, “Camera compensation using a feature projection matrix for person reidentification,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 8, pp. 1350–1361, 2014.

[23] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: Synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, no. 1, pp. 321–357, 2002.

[24] J. Gentle, *Computational Statistics*. UK: Springer, 2009.

[25] G. Schwarz, “Estimating the dimension of a model,” *Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[26] H. Liu, S. Gong, C. Loy, and X. Lin, “Person re-identification: What features are important?” in *First International Workshop on Re-Identification*, Firenze, Italy, Oct 2012.

[27] M. Enzweiler and D. M. Gavrila, “Monocular pedestrian detection: Survey and experiments,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2179 – 2195, 2009.